Constrained Optimization I

Nipun Batra

IIT Gandhinagar

July 29, 2025

Lagrangian and Duality

Nipun Batra

June 28, 2020

IIT Gandhinagar

Lectures heavily inspired by the Maths for Machine learning book

Minimax inequality
 states:max_y min_x q(x, y) ≤ min_x max_y q(x, y)

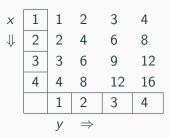
- Minimax inequality
 states: max_y min_x q(x, y) ≤ min_x max_y q(x, y)
- We first prove For all $x, y = \min_{x} q(x, y) \leq \max_{y} q(x, y)$

• Let us choose q(x, y) = xy

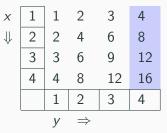
- Let us choose q(x, y) = xy
- Let us first find $\max_{y} q(x, y)$

- Let us choose q(x, y) = xy
- Let us first find $\max_{y} q(x, y)$

- Let us choose q(x, y) = xy
- Let us first find $\max_{\mathbf{y}} q(\mathbf{x}, \mathbf{y})$



• For each value of x, we find y that maximizes q(x,y)



- For each value of x, we find y that maximizes q(x,y)
- y = 4 maximizes $q(x, y) \forall x$

X	1	1	2	3	4
\downarrow	2	2	4	6	8
	3	3	6	9	12
	4	4	8	12	16
		1	2	3	4
		У	\Rightarrow		

• For each value of y, we find x that minimizes q(x, y)

X	1	1	2	3	4
\downarrow	2	2	4	6	8
	3	3	6	9	12
	4	4	8	12	16
		1	2	3	4
		у	\Rightarrow		

- For each value of y, we find x that minimizes q(x, y)
- x = 1 minimizes $q(x, y) \forall y$

X	1	1	2	3	4
\downarrow	2	2	4	6	8
	3	3	6	9	12
	4	4	8	12	16
		1	2	3	4
		у	\Rightarrow		

• We just showed For all $x, y = \min_{x} q(x, y) \leq \max_{y} q(x, y)$

X	1	1	2	3	4
 	2	2	4	6	8
	3	3	6	9	12
	4	4	8	12	16
		1	2	3	4
		У	\Rightarrow		

- We just showed For all $x, y = \min_{x} q(x, y) \leqslant \max_{y} q(x, y)$
- The equality occurs at x = 1, y = 4

X	1	1	2	3	4
\Downarrow	2	2	4	6	8
	3	3	6	9	12
	4	4	8	12	16
		1	2	3	4
		у	\Rightarrow		

• Let us now find $\max_{y} \min_{x} q(x, y)$

X	1	1	2	3	4
\downarrow	2	2	4	6	8
	3	3	6	9	12
	4	4	8	12	16
		1	2	3	4
		У	\Rightarrow		

• Similarly, let us now find $\min_{x} \max_{y} q(x, y)$

X	1	1	2	3	4
\Downarrow	2	2	4	6	8
	3	3	6	9	12
	4	4	8	12	16
		1	2	3	4
		У	\Rightarrow		

- Similarly, let us now find $\min_{x} \max_{y} q(x, y)$
- We can thus see our Minimax inequality $\max_{\mathbf{y}} \min_{\mathbf{x}} q(\mathbf{x}, \mathbf{y}) \leqslant \min_{\mathbf{x}} \max_{\mathbf{y}} q(\mathbf{x}, \mathbf{y})$

X	1	1	2	3	4
\Downarrow	2	2	4	6	8
	3	3	6	9	12
	4	4	8	12	16
		1	2	3	4
		У	\Rightarrow		

Our problem is of the form

$$\min_{\mathbf{x}} f(\mathbf{x})$$

subject to $g_i(\mathbf{x}) \leqslant 0$ for all $i = 1, \dots, m$

Our problem is of the form

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$
 subject to $g_i(\boldsymbol{x}) \leqslant 0$ for all $i=1,\ldots,m$

Idea: Convert constrained problem to an unconstrained problem

$$J(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^{m} \mathbf{1} \left(g_i(\mathbf{x}) \right)$$

Our problem is of the form

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$
 subject to $g_i(\boldsymbol{x}) \leqslant 0$ for all $i=1,\ldots,m$

Idea: Convert constrained problem to an unconstrained problem

$$J(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^{m} \mathbf{1}(g_i(\mathbf{x}))$$

where 1(z) is an infinite step function

$$\mathbf{1}(z) = \left\{ egin{array}{ll} 0 & ext{if } z \leqslant 0 \ \infty & ext{otherwise} \end{array}
ight.$$

Our problem is of the form

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$
 subject to $g_i(\boldsymbol{x}) \leqslant 0$ for all $i=1,\ldots,m$

Idea: Convert constrained problem to an unconstrained problem

$$J(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^{m} \mathbf{1} \left(g_i(\mathbf{x}) \right)$$

where 1(z) is an infinite step function

$$\mathbf{1}(z) = \begin{cases} 0 & \text{if } z \leqslant 0 \\ \infty & \text{otherwise} \end{cases}$$

This would give infinte penalty if constraint is not satisfied. But, this formulation is hard to solve too.

Idea: Introduce Lagrange multipliers $(\lambda_i \ge 0)$ to "approximate" J(x)

$$\mathfrak{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x})$$

Idea: Introduce Lagrange multipliers ($\lambda_i \geq 0$) to "approximate" J(x)

$$\mathfrak{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x})$$

What is the relationship between $\mathfrak{L}(x,\lambda)$ and J(x) given $\lambda_i \geq 0$?



Idea: Introduce Lagrange multipliers $(\lambda_i \ge 0)$ to "approximate" J(x)

$$\mathfrak{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x})$$

What is the relationship between $\mathfrak{L}(\mathbf{x}, \lambda)$ and $J(\mathbf{x})$ given $\lambda_i \geq 0$?

When $\lambda \geqslant 0$, the Lagrangian $\mathcal{L}(x,\lambda)$ is a lower bound of J(x). Hence, the maximum of $\mathfrak{L}(x,\lambda)$ with respect to λ is

$$J(\mathbf{x}) = \max_{\mathbf{\lambda} \geqslant 0} \mathfrak{L}(\mathbf{x}, \mathbf{\lambda})$$





$$J(\mathbf{x}) = \max_{\mathbf{\lambda} \geqslant 0} \mathfrak{L}(\mathbf{x}, \mathbf{\lambda})$$

$$J(\mathbf{x}) = \max_{\mathbf{\lambda} \geqslant 0} \mathfrak{L}(\mathbf{x}, \mathbf{\lambda})$$

But, our original problem was minimizing J(x), which is equivalent to:

$$\min_{oldsymbol{x} \in \mathbb{R}^d} \max_{oldsymbol{\lambda} \geqslant oldsymbol{0}} \mathfrak{L}(oldsymbol{x}, oldsymbol{\lambda})$$

$$J(\mathbf{x}) = \max_{\mathbf{\lambda} \geqslant 0} \mathfrak{L}(\mathbf{x}, \mathbf{\lambda})$$

But, our original problem was minimizing J(x), which is equivalent to:

$$\min_{oldsymbol{x} \in \mathbb{R}^d} \max_{oldsymbol{\lambda} \geqslant oldsymbol{0}} \mathfrak{L}(oldsymbol{x}, oldsymbol{\lambda})$$

Using the Minimax inequality, we can write:

$$J(\mathbf{x}) = \max_{\mathbf{\lambda} \geqslant 0} \mathfrak{L}(\mathbf{x}, \mathbf{\lambda})$$

But, our original problem was minimizing J(x), which is equivalent to:

$$\min_{oldsymbol{x} \in \mathbb{R}^d} \max_{oldsymbol{\lambda} \geqslant oldsymbol{0}} \mathfrak{L}(oldsymbol{x}, oldsymbol{\lambda})$$

Using the Minimax inequality, we can write:

$$\min_{\boldsymbol{x} \in \mathbb{R}^d} \max_{\boldsymbol{\lambda} \geqslant \boldsymbol{0}} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \geqslant \max_{\boldsymbol{\lambda} \geqslant \boldsymbol{0}} \min_{\boldsymbol{x} \in \mathbb{R}^d} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

$$J(\mathbf{x}) = \max_{\mathbf{\lambda} \geqslant 0} \mathfrak{L}(\mathbf{x}, \mathbf{\lambda})$$

But, our original problem was minimizing J(x), which is equivalent to:

$$\min_{oldsymbol{x} \in \mathbb{R}^d} \max_{oldsymbol{\lambda} \geqslant \mathbf{0}} \mathfrak{L}(oldsymbol{x}, oldsymbol{\lambda})$$

Using the Minimax inequality, we can write:

$$\min_{\boldsymbol{x} \in \mathbb{R}^d} \max_{\boldsymbol{\lambda} \geqslant \boldsymbol{0}} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \geqslant \max_{\boldsymbol{\lambda} \geqslant \boldsymbol{0}} \min_{\boldsymbol{x} \in \mathbb{R}^d} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

We can write the dual objective as a function of λ as

$$\mathfrak{D}(\boldsymbol{\lambda}) = \min_{\boldsymbol{x} \in \mathbb{R}^d} \mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$



• Primal objective:

$$\min_{m{x}} f(m{x})$$
 subject to $g_i(m{x}) \leqslant 0$ for all $i=1,\ldots,m$

• Primal objective:

$$\min_{\mathbf{x}} f(\mathbf{x})$$

subject to $g_i(\mathbf{x}) \leqslant 0$ for all $i = 1, \dots, m$

ullet Or, primal objective $=J(x)\geq \max_{oldsymbol{\lambda}}\mathfrak{D}(oldsymbol{\lambda})$

Primal objective:

$$\min_{\mathbf{x}} f(\mathbf{x})$$

subject to $g_i(\mathbf{x}) \leq 0$ for all $i = 1, ..., m$

- Or, primal objective = $J(x) \ge \max_{\lambda} \mathfrak{D}(\lambda)$
- Or, primal objective (in terms of x) \geq dual objective (in terms of λ)

• Primal objective:

$$\min_{\mathbf{x}} f(\mathbf{x})$$

subject to $g_i(\mathbf{x}) \leqslant 0$ for all $i = 1, \dots, m$

- Or, primal objective = $J(x) \ge \max_{\lambda} \mathfrak{D}(\lambda)$
- Or, primal objective (in terms of x) \geq dual objective (in terms of λ)
- For SVM like formulations, primal objective is the same as dual objective (strong duality)

Primal objective:

$$\min_{\mathbf{x}} f(\mathbf{x})$$

subject to $g_i(\mathbf{x}) \leqslant 0$ for all $i = 1, \dots, m$

- Or, primal objective = $J(x) \ge \max_{\lambda} \mathfrak{D}(\lambda)$
- Or, primal objective (in terms of x) ≥ dual objective (in terms of λ)
- For SVM like formulations, primal objective is the same as dual objective (strong duality)
- For some problems, there is a "daulity-gap" between the two objectives